ExpoBlend: Information preserving exposure blending based on normalized log-domain entropy

Article *in* Computers & Graphics · April 2014

DOI: 10.1016/j.cag.2013.10.001

1 author:

Neil D. B. Bruce
University of Manitoba
46 PUBLICATIONS **1,103** CITATIONS

SEE PROFILE

Computers & Graphics 39 (2014) 12–23

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on Graphics Interface

ExpoBlend: Information preserving exposure blending based on normalized log-domain entropy

Neil D.B. Bruce *

University of Manitoba, Department of Computer Science, 66 Chancellors Circle, E2-408 EITC, Winnipeg, MB, Canada R3T 2N2

ARTICLE INFO

Article history:
Received 10 June 2013
Received in revised form
18 September 2013
Accepted 2 October 2013
Available online 17 October 2013

Keywords:
Exposure blending
Illumination
Reflectance
Computational photography
Tone-mapping
Entropy

ABSTRACT

In this paper, we present a solution to the problem of dynamic range compression from multiple exposures called ExpoBlend that operates in the absence of raw format images, relative or absolute exposure values, camera response functions, or known irradiance. This is achieved in relatively simplistic fashion by merging image content across provided exposures. The proposed algorithm is directed at making visible any contrast appearing across a dynamic range that exceeds display or printing capabilities through high dynamic range (HDR) compression while preserving the nature of the image structure and detail, lighting, and avoiding introducing discontinuities in illumination or image artifacts. In addition, ExpoBlend allows scaling subject to a single parameter that elicits a trade-off between the impact of illumination and fine detail in the merged result. The strategy applied appeals to an information maximization strategy wherein the local entropy evident in each exposure is computed subject to a logarithmic compression of intensities, and employs cross-exposure normalization of entropy that implies a fusion strategy based on relative entropy across exposures in combination with a soft-maximum operation.

1. Introduction

Recent efforts have produced many innovations in the field of computational photography ranging from algorithmic efforts to provide augmented depictions of content captured by visual sensors, to innovative contributions to the nature of the hardware, optics and electronics involved in capture and display. While recently there has been a gradual shift in emphasis to innovations in hardware technology, a problem of significant focus in the past decade has been in finding means of capturing visual content in a manner that carries a much higher dynamic range. The most common means of achieving this has been in capturing a number of bracketed exposures, which are subsequently combined subject to a known or estimated radiometric relationship between irradiance and intensity, capturing imagery across a higher dynamic range than a single exposure might present.

While this presents a rich set of data for analysis and display, it presents the problem of finding means of displaying HDR data in a manner suitable for lower dynamic range display technology. Existing research efforts present a multitude of distinct strategies devoted to the problem of compressing the range of visual imagery comprising a high dynamic range, down to a level suitable for this purpose. The exact goals of performing this operation are somewhat varied, with some efforts concerned primarily with producing images that are of a pleasing qualitative nature, providing realism while maximally preserving detail, or producing cartoon-like or detail augmented output.

While these efforts serve as important tools in the arsenal of capture and display technology, the overall process can be laborious and tends to call for hardware that allows a minimum level of control and stability to derive high quality radiometric response mappings from distinct exposures, or estimation of irradiance from captured imagery. In some instances, hardware allows for capture of reasonably high quality imagery (e.g. recent HD webcams), but does not carry the same properties of allowing absolute (or relative) exposure to be known. Proprietary hardware in conjunction with proprietary drivers often results in conditions where the broad range of approaches for estimating the radiometric response curves are unsuitable, producing either unstable or inaccurate results. Moreover, even in the presence of reasonably well behaved hardware for image capture, a means of combining independent exposures without the need to resort to careful measurements or calibration procedures is of value in providing the ability to present captured content on a standard display in a relatively quick and simple fashion.

To this end, we present an algorithm for combining content across a number of given exposures with unknown image parameters (e.g. exposure) in a manner that appeals to best capturing

* Tel.: +1 204 474 7313.
E-mail addresses: bruce@cs.umanitoba.ca, neilbruce0@gmail.com
URL: http://www.cs.umanitoba.ca/~bruce/

0097-8493/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cag.2013.10.001

the information present in independent exposures. This is done in appealing to the relative local entropy across images, with a single parameter inducing a non-linearity which presents a trade-off between detail that is visible, and the smoothness of the fused result. In carrying out this exercise, we demonstrate that very good results may be had based on a handful of exposures in a fashion that is less labor intensive than existing strategies and arguably having the capability of providing a better end result. We also discuss various strategies to achieve exposure fusion, and contrast them against the current proposal in establishing the benefits of ExpoBlend relative to alternative methods. This builds on our prior work [1] in further understanding the impetus for behavior that is observed, and in connecting the work to other canonical computational operations (specifically the softmax activation function that appears in neural networks). This is further bolstered by additional experimental results across an array of image types, along with discussion of qualitative behavior, quantitative results, and challenges associated with quantifying performance.

A description of some of the more prominent exposure blending algorithms that appear in the literature is presented in Section 2 along with a selection of traditional efforts in tone mapping. In Section 3, we present the details of the ExpoBlend algorithm followed by a detailed discussion of motivating principles and observations relating to the algorithm in Section 4. Section 5 describes the methodology for evaluation followed by comparison of results from the ExpoBlend method against a number of commonly used algorithms in Section 6. Finally, in Section 7, we present discussion relating to the conditions under which one may expect very high quality results, the relationship between images and parameter adjustment, and some natural extensions of ExpoBlend that may allow for further improvements, or use cases.

2. Background

Historically, many efforts toward dynamic range compression, or exposure blending appeal to recovery of the relationship between irradiance, and intensity values to linearize said relationship. This allows for subsequent fusion of image content appearing across multiple exposures through strategies that employ tone-mapping or image blending. It is important to note that the literature on tone-mapping is considerable, and the algorithms discussed represent a cross-section of some of the more prominent and formative efforts in this domain. An alternate set of approaches attempts to directly compose a combined image in the absence of determination of a camera response function, irradiance or other photometric parameters. In the following, we discuss a number of representative, and formative efforts in tone-mapping and exposure blending. Many of these form the basis for comparative results that appear in evaluation results presented in the paper.

2.1. Tone-mapping operations

Fattal et al. [2] propose a method that involves compression in the gradient domain. This is accomplished in solving the Poisson equation for the compressed gradient field. In varying the amount of compression, one is able to modulate the level of detail present in the resulting output. Reinhard et al. [3] propose a method inspired by Ansel Adams zone system that appeals to strategies that have history in wet photography. This includes considering brightness, contrast, and also viewing conditions and employs a strategy analogous to the dodge and burn techniques used in classic photography. Durand and Dorsey [4] propose a method based on bilateral filtering that typically results in smoothing an image without corrupting edge content. In this manner, one may separate the detail content from broader intensity variation. The base-layer low-detail image is used to modulate the high-detail image to produce a tone-mapped version in a multiplicative manner. Drago et al. [5] use logarithmic compression of the dynamic range of the HDR content. This is done in adapting the base of the logarithm in a manner that preserves local detail. This results in stronger preservation of detail in darker regions, and greater compression in bright regions. Reinhard and Devlin [6] present a distinct approach from the earlier Reinhard et al. work that appeals to photoreceptor physiology in the human visual system. This includes adaptation of cone responses as a function of intensity, and minimum and maximum log luminance ratios towards a characterization that adapts content based on adjustment of contrast and intensity, while also performing chromatic and light adaptation in a manner that is aligned with the behavior of the human visual system. Mantiuk et al. [7] also employ a perceptually motivated strategy based on contrast mapping and contrast equalization appealing to physiological measurements tied to factors such as the Weber contrast, and logarithmic ratios in considering a multi-scale representation within a Gaussian pyramid.

This initial effort is further expanded on [8], also including consideration of typical LCD display properties, and combined enhancement and estimated visibility of features based on the human visual system. This is done in a manner that determines a minimum error tone-mapped image subject to both the display model, and enhancement of the image cast into a representation consistent with the human visual system. Error in this regard appeals to a minimum distortion metric in determining the resulting output.

2.2. Image blending

There exist a handful of methods for achieving a similar effect to tone-mapping via fusion of exposures that have been proposed in recent years. Each of these appeal to different local image properties. The most influential of efforts in this domain, and those that are most closely related to the proposal at hand are summarized below.

In the “Exposure blending” or *enfuse* algorithm of Mertens et al. [9], local measures are derived based on quantitative notions of contrast, exposure and well-exposedness. This is perhaps the most widely used algorithm for blending content across exposures. Given the possibility of introducing local discontinuities in intensity in blending, this operation is performed on a Laplacian pyramid. This is a very sensible strategy for overcoming this issue, however it does imply that relatively low spatial frequency intensity variations can serve to corrupt detail in what may be highly localized regions of intensity variation, coupled with less localized but more severe variations in irradiance.

There is also an alternative for blending exposures based on entropy presented by Goshtasby [10]. In this work, the image is divided into a grid of local windows. The entropy of these windows is computed, and the best exposure is assigned to each window based on entropy. Fusion is then achieved by interpolation based on the exposures assigned to each window. This strategy again suffers from non-locality and also does not allow precise control over the trade-off between aggressively selecting content from a small set of exposures, while simultaneously allowing for broader more graded cross-exposure combination. In addition, the hard selection of particular exposures from single images, and spatially dispersed image samples present additional limitations on overall performance and flexibility in adapting the nature of algorithm output.

All of these methods share the same challenge as the ExpoBlend algorithm: Given what may be "jumps" in the local contrast from one exposure to another within relatively unstructured regions of the scene, some means of creating a smooth transition between the best exposures for any local region is necessary. In the *enfuse* method [9] this is accomplished via Laplacian pyramids, implying a multi-scale fusion operation for which a smooth transition may be ascertained. In the entropy motivated method of Goshtasby [10], the discrete separated image partitions are blended based on a piece-wise linear blending operation. In the current work, this smoothness is achieved via the local log-domain normalized entropy coupled with a soft-max non-linearity, ensuring that the contribution across exposures always sums to 1. This allows for one or several exposures to dominate on the basis of how evenly spread the local entropy score appears across exposures. This has the advantage that for areas in which the detail is highly localized, the combination is not corrupted by contrast present at an overly large scale or tied to very low spatial frequency variations in image content. This allows for isolated detail that is very localized both in space and within few or singular exposures to produce better contrast than what is achievable based on traditional tone-mapping operators, and alternative blending strategies.

3. ExpoBlend: normalized entropy-based blending

The following presents the algorithm that is employed in order to combine a distinct set of exposures to produce a combined result suitable for LDR display, while also allowing control between detail and smooth spatial variation in output on the basis of a single parameter. This is applicable to the original exposures, and as such, neither linearization or gamma correction of the captured images are required for the algorithm to produce desirable output.

Given a set of images $\{I_k\}$ with the same viewpoint taken at different exposures k with $k=1\ldots N$, the exposure blended image is produced as follows:

For a given pixel location i, j across the image exposures k, with $I_k(i,j)\in[0,1]$:

$$I_k^{\log}(i,j)=\log(1+I_k(i,j)) \tag{1}$$

The local entropy $H_k^{\log}(i,j)$ of $I_k^{\log}(i,j)$ is computed within a flat local circular support $S=\{I_k^{\log}(x,y)\}$ where

$$\sqrt{(i-x)^2+(j-y)^2}<R \tag{2}$$

with R a constant defining the radius of the support region. Let $H_k^{norm}(i,j)$ represent the cross exposure normalized log-scale local entropy given by

$$H_k^{norm}(i,j)=\frac{H_k^{\log}(i,j)}{\sum_{k=1}^{N}H_k^{\log}(i,j)} \tag{3}$$

This is subsequently subjected to a non-linear adjustment of values, that with normalization implies $H_k^{norm}(i,j)\in[0,1]\ \forall k$. This is subject to the equation:

$$H_k^{nlnorm}(i,j)=e^{\beta H_k^{norm}(i,j)} \tag{4}$$

where β is a parameter that controls the extent to which the weighting in fusion of exposures is spread across distinct exposures.

Finally, the values of $H_k^{nlnorm}(i,j)$ are subsequently normalized to form the weights for each exposure k at i, j:

$$H_k^{weights}(i,j)=\frac{H_k^{nlnorm}(i,j)}{\sum_{k=1}^{N}H_k^{nlnorm}(i,j)} \tag{5}$$

Given the resulting weights, the values at i, j in the blended image are given by

$$I^{blended}(i,j)=e^{\sum_{k=1}^{N}H_k^{weights}(i,j)I_k^{\log}(i,j)} \tag{6}$$

Pseudo-code for the ExpoBlend algorithm follows for additional clarity. In the instance that there is minor variation in viewpoint subject to any camera movement, alignment may be required as is the case with all existing approaches to the problem in question and the following assumes aligned images. In the following I(k) refers to I_k above.

```
for k from 1 to N:
  Convert I(k) to double precision
  I(k)=I(k)/255
  logI(k)=log(1+I(K))

totalE1=0
totalE2=0

for k from 1 to N:
  E(k)=Local entropy of logI(k)
  (based on local M×M windows)
  totalE1=totalE1+E(k)

for k from 1 to N:
  E(k)=E(k)/totalE1
  E(k)=exp(βE(k))
  totalE2=totalE2+E(k)

for k from 1 to N:
  E(k)=E(k)/totalE2
  finalBlended=finalBlended+E(k)°logI(k)

finalBlended=exp(finalBlended)
```

The ° symbol corresponds to the standard notation for the Hadamard product (pointwise multiplication)

4. ExpoBlend: motivating principles and observations

In the preceding description of the algorithm, there are a few nuances to the computation that is performed that are simplistic in their nature, but essential to achieving high quality results from the proposed algorithm. The design decisions, and related observations tied to the computation performed are summarized in the following.

In determining local entropy within the image, each individual exposure is first subjected to a compressive logarithmic non-linearity (Eq. (1)). One can point to at least a few sources that shed light on the potential benefit of this consideration. Perhaps the most direct motivation for log-domain processing of exposures comes from the motivating principles that inspire homomorphic filtering [11]. Given that illumination varies across the image at a relatively slow rate when compared with reflectance, which may give rise to more local variation, log-domain measurement of local entropy will tend to discount the impact of variation in illumination across any given exposure, and provide a measure of information more strongly dependent on reflectance. This interpretation treats local variation in illumination as multiplicative noise. An additional connection of interest is derived from consideration of strategies for determining *intrinsic images*, involving the separation of an image into reflectance, shading, and specular components [12]. Some of the more successful efforts in this domain appeal to the notion that local gradients in the log-domain provide a diagnostic for variation in reflectance. In the current work this behavior was found to provide benefits in preserving contrast and

detail present in blending exposures, while also having the ability to elicit a trade-off between the impact of local illumination and reflectance on the blended result (subject to the parameter β).

Given the relatively small differences in entropy when measured on a logarithmic scale, the weighted contribution of exposures is subject to an exponential weighting (Eq. (4)). At first glance this may seem to add overhead where it is not required, however the entropy computation being subject to summation over a local region implies a narrow range of cross-exposure differences in entropy in some cases. The exponential operation serves the role of constraining the β parameter to reside within a smaller and more predictable range of values.

Perhaps the most paramount operation in this pipeline is the cross exposure normalization, followed by a non-linearity, followed by normalization (Eqs. 3–5 respectively). This normalization, non-linearity, normalization is central to achieving locality, smoothness, and allowing scaling between extremes of continuity in the representation of more global illumination or smoothness in general, and level of detail that is visible. It is interesting to note that the three steps involved imply consideration of relative rather than absolute entropy across exposures (Eq. (3)), followed by combination through an operation equivalent to the softmax activation function that appears in neural networks [13] (Eqs. 4 and 5) with the normalized log-entropy providing connection weights.

One scenario where one may observe less than ideal blending arises in the complete absence of any local image structure (e.g. a large blank region of wall). Typically for high quality imagery, even for a relatively local region subtle variations corresponding to signal in the image are sufficient to maintain a highly local spatial extent for quantifying local entropy. Failing this, experimentation suggests that this scenario can be resolved by maintaining a lower value for β, or expanding the locality of the filter aperture over which entropy is measured. In practice, the latter option seems to have a relatively small impact on the ability to retain detail within the image without any contingencies, especially given that the β value can be increased to compensate for loss of contrast. In the case of the *memorial* comparison (Fig. 1) and existing HDR to LDR fusions (Fig. 8), blended exposures correspond to a value of $R=11$. For all other cases, $R=59$ has been used, for the stated limitation that very flat regions can become darkened with a highly local support. Note that this expansion of the filter aperture does not apparently have a strong impact on the ability to preserve highly local detail given the trade-off between filter aperture and scaling of algorithm behavior subject to β.

5. Evaluation methodology

In order to evaluate the ExpoBlend algorithm, we provide a qualitative comparison against an array of some of the most widely used and cited algorithms in the literature discussed prior. This is coupled with quantitative results for images considered, while also discussing challenges in quantifying performance in HDR to LDR conversion. Algorithms considered include both tone-mapping algorithms and image blending algorithms. Given the large number of algorithms that have been proposed, the focus of the evaluation is on more prominent and recent algorithms appearing in the literature. This includes a detailed examination of perhaps the most popular image that appears in this literature (the *memorial* image) and additionally, a comparison against a variety of algorithms implemented by Čadík based on an image with a very large dynamic range for which original source images are available. In the case of tone-mapping, seven different tone-mapping algorithms have been considered in the detailed *memorial* evaluation. In addition, the classic *memorial* image presents a very similar set of exposures to those employed in the "Exposure Blending" or *enfuse* paper of Mertens et al. [9] (differing in the viewpoint of the photos, but very similar in the nature of the examples considered presenting a good basis for comparison). For this reason, the examples produced from Mertens et al.'s algorithm are included for comparison, as well as output for the blending approach of Goshtasby [10] based on the original *memorial* image. These are perhaps the most prominent examples in the literature that appeal to a direct blending approach. Although software for the *enfuse* algorithm is available, given that the output of the *enfuse* algorithm is dependent on the parameter choices, lack of source images for those that appear in Mertens et al.'s results [9] and the similarity of Mertens et al.'s example of output to the classic *memorial* image, this was deemed to provide the best impartial basis for comparison. In particular, this example is sufficient to demonstrate what are significant differences among the high end of the dynamic range.

Tone-mapping output for the *memorial* image is based on implementations provided by the *pfstools* software library [18]. Many of the original papers indicate that post-processing adjustment of the gamma of the output is necessary, and attempts have been made in each instance to select a level of gamma adjustment that is most flattering in revealing the extent of the compressed range for each algorithm (output also similar to results presented elsewhere). This is true of all of the algorithms presented with one exception [8], which does not require post-processing adjustment of gamma.

6. Results

ExpoBlend has been applied to images commonly used in examining tone-mapping and exposure blending, and a comparison of the output of ExpoBlend to a number of classic and popular algorithms from the tone-mapping and exposure blending literature is presented. Despite operating on a rather crude representation relative to the richer set of information available to these algorithms, that generally employ knowledge of irradiance or photometric curves by design (and rely on .hdr or .exr format images), we demonstrate that very good results may be achieved without going to such lengths.

6.1. Comparison with tone-mapping operators

As a basis for comparing the output of ExpoBlend with some of the classic tone-mapping strategies, a comparison against the seven tone-mapping operators described in Section 2 is presented. This comparison appears in Fig. 1. In Fig. 1, the output of each of the algorithms on the widely used *memorial* image [16] is shown. It is apparent, given the focus of discussion on localization, that there is a trade-off between creating a smooth transition of illumination in blending exposures, and retaining detail. This is especially true at extremes in irradiance appearing in HDR images such as exceptionally bright or dark regions. A close-up of the window region from the *memorial* image as demonstrated in Fig. 2 reveals that one is better able to preserve locality of detail in such regions.

6.2. Effect of the non-linearity

It is important to consider that the results presented yield a relatively continuous variation in the image, avoiding significant jumps in local luminance without appealing to either multi-scale strategies, or any explicit spatial interpolation. This is achieved via the normalization-softmax pipeline. One also has the advantage that the trade-off between the impact of local variation in illumination

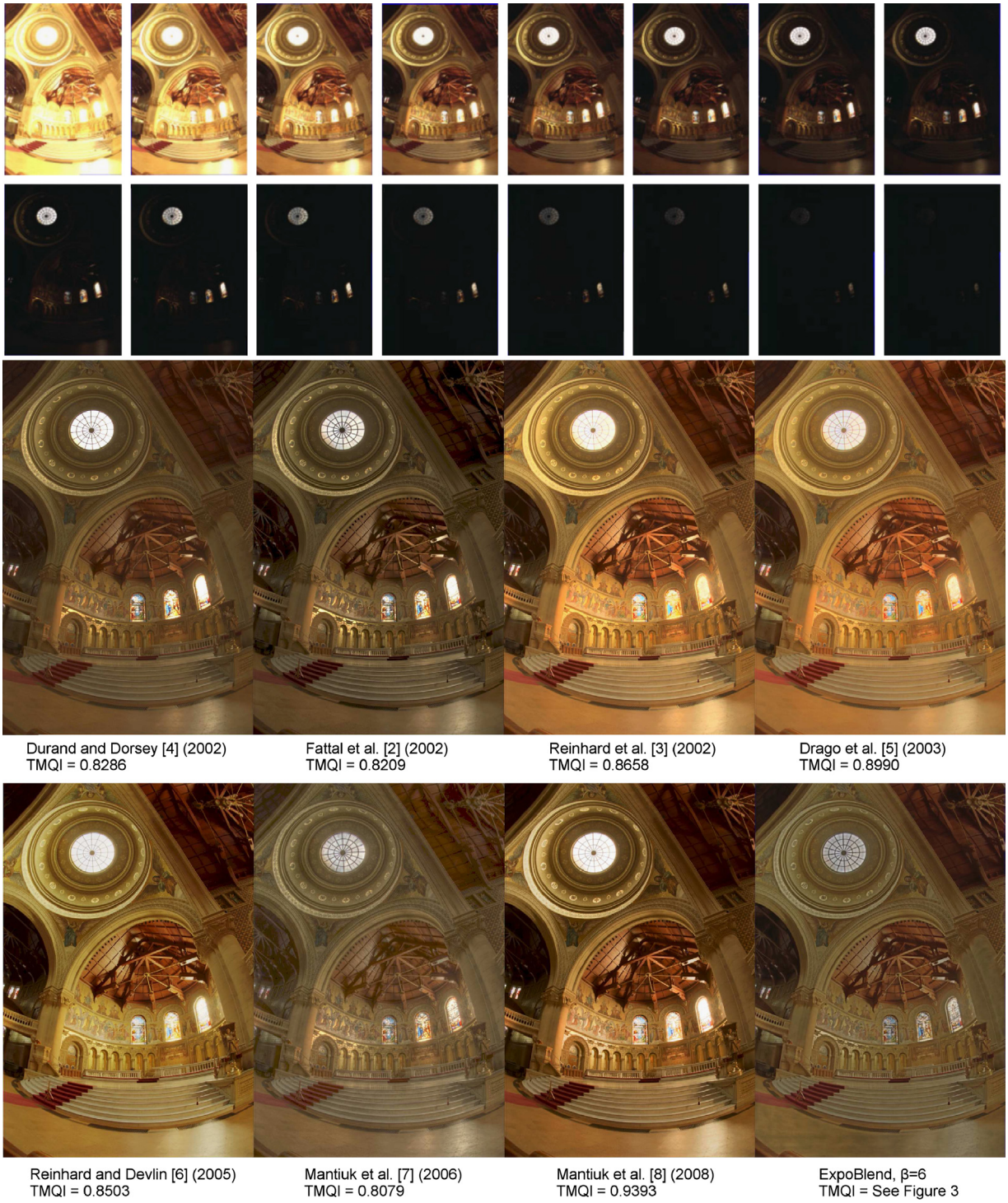

Fig. 1. A comparison of algorithms on the *memorial* image exposures. Annotations below each image refer to the algorithm that produced the LDR result. The TMQI score for each image is also given. Note also that for ExpoBlend, one can vary the output in a manner that implies a trade-off between smoothness of illumination and detail visible. See Fig. 3 for a depiction of this behavior.

and scene structure may be manipulated via a single parameter which allows scaling between strong spatial continuity of illumination, versus a heavy emphasis on exposing detail towards a less naturalistic high contrast output. Fig. 3 demonstrates the role of this parameter in considering the *memorial* image. For typical cases with a small number of exposures, appropriate values for this single parameter lie in the range of 0.5–6. If one has more exposures available (or complete HDR content – as discussed in what follows), one can be even more aggressive in the choice of this parameter, and pushing this value as high as 10–15 can produce naturalistic images while exposing all available detail. To produce an image that is less *realistic*, but maximizes the edge content and contrast visible, one can employ β values as high as 50–200. These ranges are quite consistent across images, and choosing a best guess based on desired output is quite easy to achieve. Moreover, as the parameter space involves only a single parameter, it is easy to adjust this to achieve output that is consistent with what one desires as output whether it be a high-fidelity representation of real-world content distributed

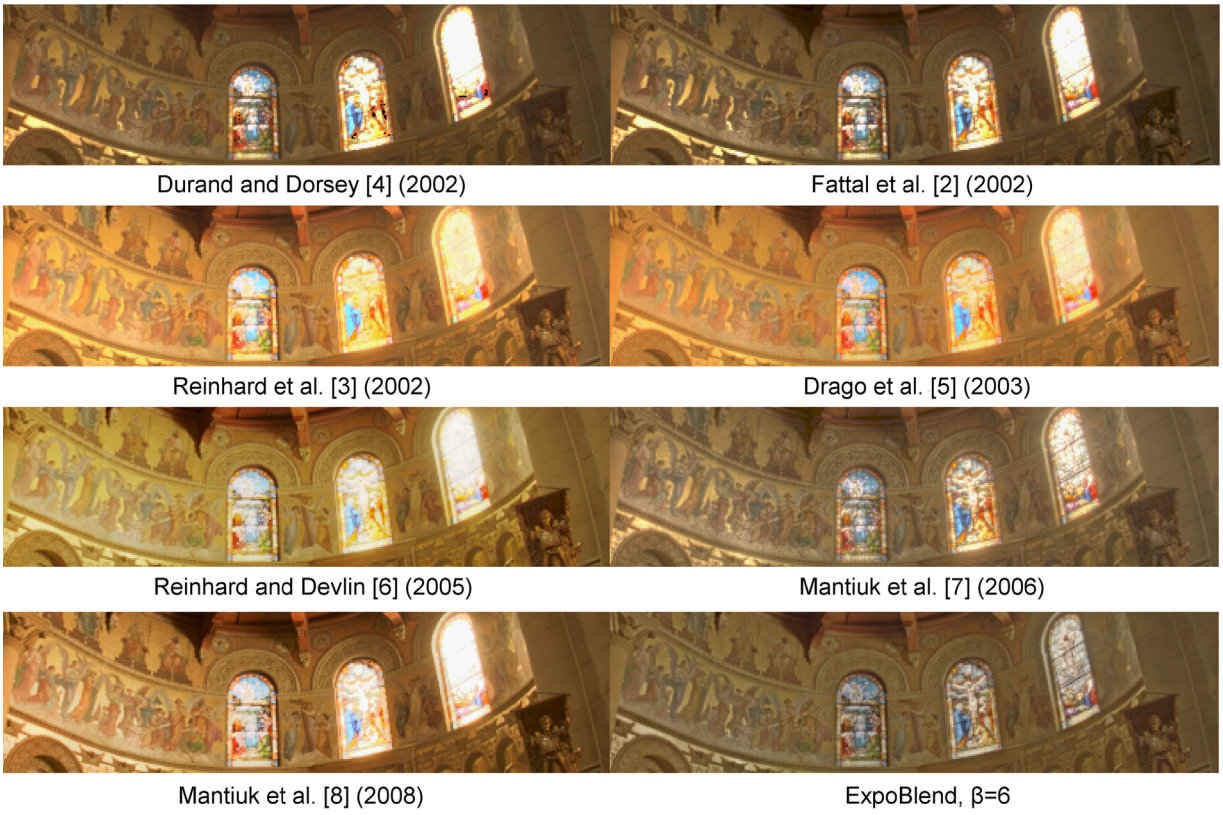

Fig. 2. Close-up of the window region based on the algorithms applied to the *memorial* image appearing in Fig. 1. Annotations below each image refer to the algorithm that produced the resulting image.

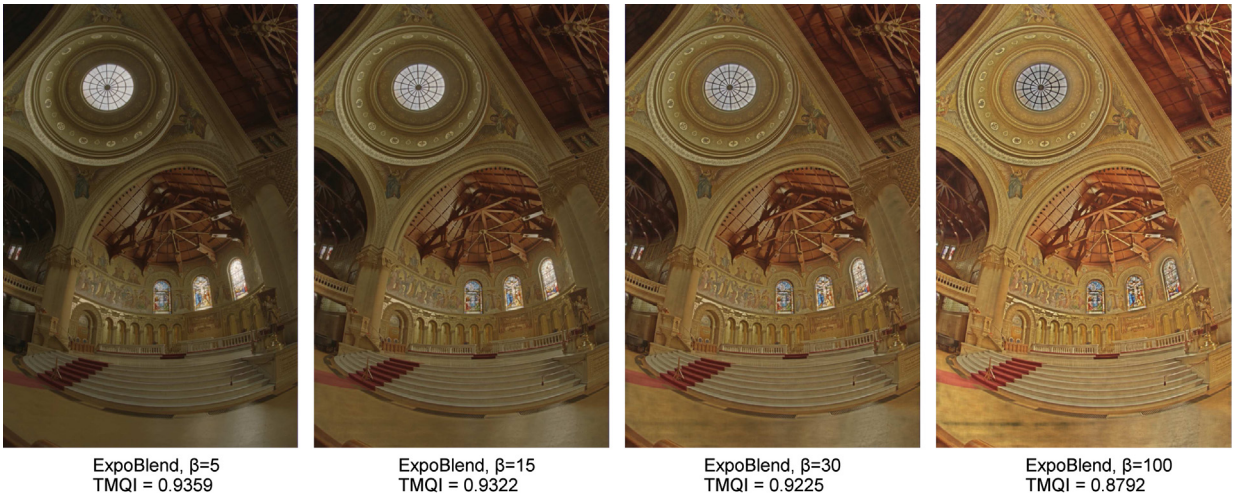

Fig. 3. A demonstration of the effects of parameter variation on smoothness and detail. Based on the 16 available exposures, one may choose more aggressive values for the value of β (5, 15, 30 and 100) respectively in this case. TMQI scores are given for each of the parameter choices for ExpoBlend on the *memorial* image.

over a high-dynamic-range, artistic, or detail emphasizing rendering of the image.

6.3. Comparison with blending operations

For a comparison with the exposure blending approach of Mertens et al., an example very similar to the *memorial* image appearing in Mertens et al.'s own work is presented [9]. Fig. 4 demonstrates the output of the algorithm for a different view but similar set of raw exposures (bottom row, second from right). One may observe that in this instance, there is also a greater degree of detail that reveals itself at the high end of the dynamic range in employing the ExpoBlend approach. Also shown is the output of Goshtasby's [10] blending approach on the classic *memorial* image (see figure annotations). Mertens et al.'s comparison to the classic techniques of Ogden et al. [14] and Burt et al. [15] are also included.

6.4. Additional comparisons

Towards providing greater breadth in the assessment of performance, two examples of algorithm performance are presented in which a very wide dynamic range is present across exposures. The first of these, presented in Fig. 5 depicts original exposures, in

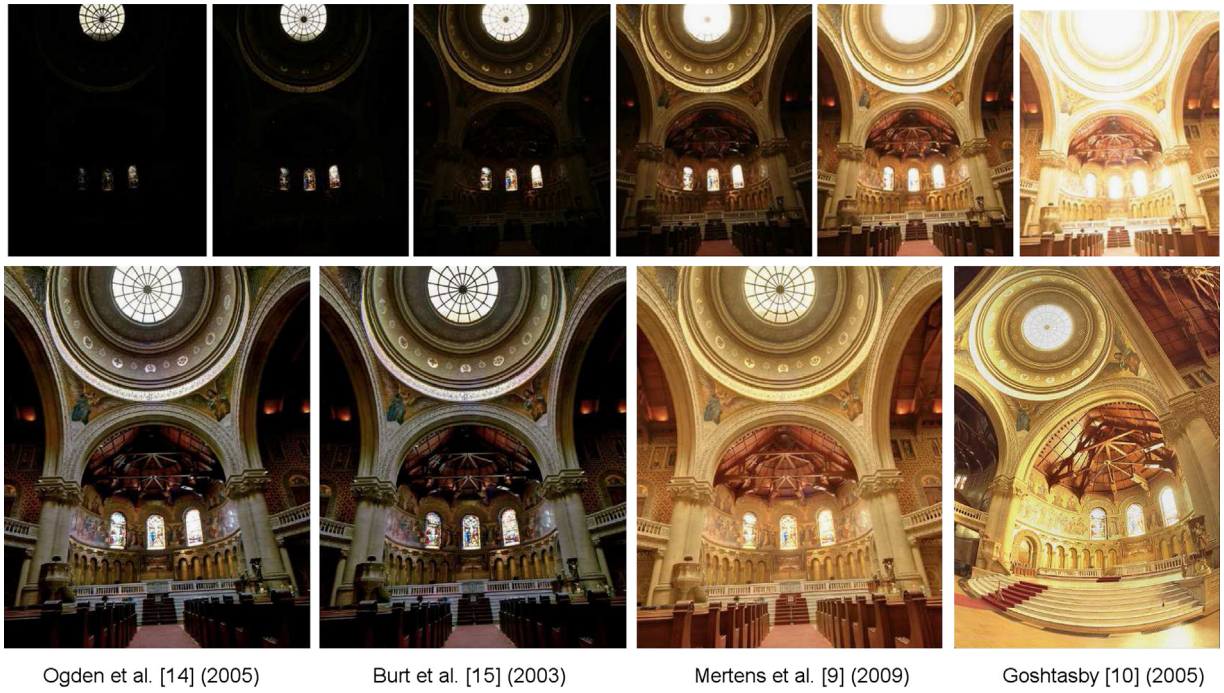

Fig. 4. An example derived from the output of the *enfuse* blending algorithm of Mertens et al. [9], who also included comparisons with Ogden et al. [14] and Burt et al. [15] for the *memorial* cathedral taken from a different viewpoint. Also included is the blended output of Goshtasby [10] on the *memorial* exposures provided by Debevec and Malik [16].

Fig. 5. An additional example of ExpoBlend fusion results for a scenario with a high-dynamic range. Exposures are shown on the top row. Bottom row shows fused output from ExpoBlend corresponding to β values of 6 and 100 respectively.

addition to blended output subject to a β value of 6 and 100 respectively. Note the ability to render HDR detail visible in a flexible manner subject to variation in the parameter β. Fig. 6 presents a comparison of output of the ExpoBlend algorithm as compared with a wide array of results from a number of algorithms implemented in the comparative study performed by Čadík et al. [17,19,20]. Some of the results shown overlap with algorithms appearing in Fig. 1, but there are also a number of additional comparisons in this case. For the precise details of these algorithms, the reader is encouraged to consult [19]. In the examples shown in Fig. 6, the algorithms and associated references appear in figure annotations. Algorithms appearing in this

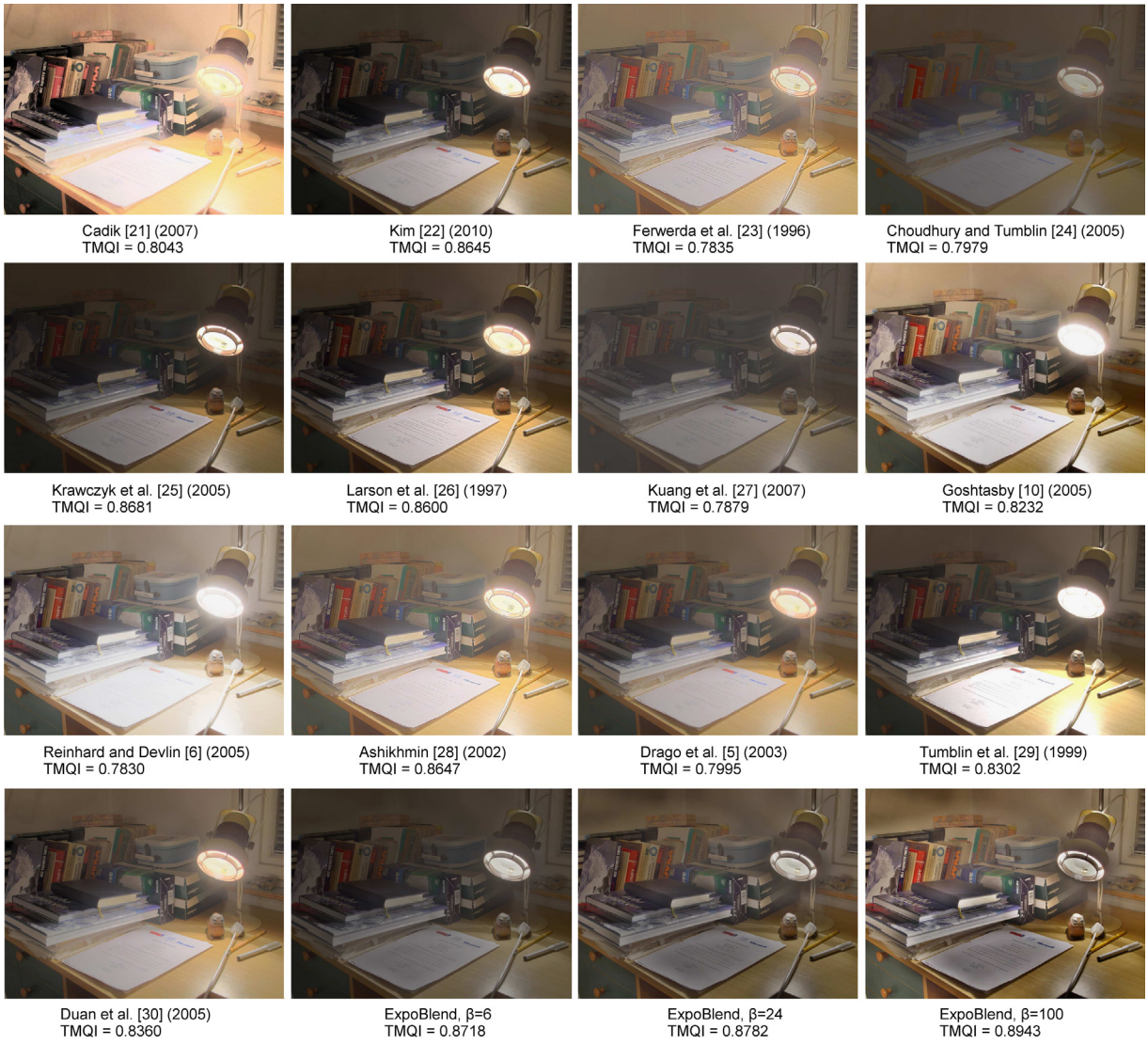

Fig. 6. Comparison of output against a broad set of tone-mapping and blending algorithms based on the exposures provided by Cadik [17]. Each image is annotated with the algorithm that produced the resulting image, and the associated TMQI score. ExpoBlend results are shown corresponding to β values of 6, 24 and 100.

comparison include those of Cădík [21], Kim [22], Ferwerda et al. [23], Choudhury and Tumblin [24], Krawczyk et al. [25], Larson et al. [26], Kuang et al. [27], Goshtasby [10], Reinhard and Devlin [6], Ashikhmin [28], Drago et al. [5], Tumblin et al. [29], Duan et al. [30], and ExpoBlend for multiple values of β. Note that the output can be driven towards an appearance subjectively reminiscent of some of the alternative algorithms subject to variation in the β value, with good preservation of detail even in the vicinity of the bright light.

6.5. Webcam image blending

Given that one of the areas where direct blending approaches stand to present the greatest impact is those for which careful principled determination of imaging properties, or irradiance prove challenging (as is the case in considering webcam and camera phone imagery), we demonstrate fusion under challenging realistic lighting conditions subject to capture from a standard webcam. The individual exposures and results of applying ExpoBlend are shown in Fig. 7. Although the algorithm succeeds in producing a desirable image in this scenario, pushing the boundaries to an extreme level of contrast in this instance was found to have the possibility of introducing darker patches among very flat regions of the image (e.g. the rear wall). Examination of exposures suggests that this is due to a greater impact of noise at low exposure levels, and in the absence of structure, these darker images can receive a stronger weight. As discussed earlier, it may be of benefit to allow the effective filter aperture over which relative entropy is computed to expand in such instances subject to a threshold corresponding to the maximal local entropy across all exposures, or the combined sum across exposures meeting a minimum threshold. Ongoing work is directed at examining this consideration towards allowing for more aggressive fusion under poor signal to noise conditions in unstructured regions of an image. In Fig. 7, the ExpoBlend result is contrasted against the best bracketed single exposure to demonstrate effectiveness for webcam imagery, and in a scenario with less drastic lighting conditions.

6.6. Existing HDR to LDR blending

Given some of the advantages that this strategy may carry over the traditional tone-mapping operations at the level of quality of the resulting fused image, one might benefit from applying this to the case where one *does* have an existing linearized HDR representation or irradiance, by extracting LDR bracketed images from

Fig. 7. A demonstration of fusion of exposures captured from a standard Logitech webcam, with $\beta=6$.

existing HDR images for blending. We have also investigated this case by extracting a larger number of bracketed regions from existing HDR format images. Given that these images are linearized, to produce a comparison more consistent with the case of employing original exposures, bracketed regions extracted have been subjected to a non-linear inverse gamma correction by a factor of 2.2. This has been performed for a number of HDR images which were then processed by the proposed ExpoBlend algorithm. Results of this operation appear in Fig. 8.

For the examples shown, the fusion of information is derived across a significant number of different images, in some cases none of which contain a range that is remotely close to capturing the complete range of content that appears in the combined LDR image, and also none of which provide access to such significant detail in intensity variation. Fig. 9 demonstrates original exposures extracted from an HDR image for the cases shown in row 2 of Fig. 8. Note the relatively sophisticated fusion of information that is produced from widely varying imagery including the range of output that can be achieved by altering a parameter that controls a single non-linearity.

6.7. *Quantitative results and judging algorithm performance*

Judging algorithm performance is a challenging problem being both subjective and dependent upon the purpose of tone-mapping or exposure blending. In some instances, one may wish to render visible all of the image content within a low dynamic range, in a fashion that is both detail preserving and provides for realism. In other instances, the end result may be aimed at having a more artistic appeal, an extreme focus on detail recovery, or an image with overblown contrast and color for visual appeal.

One natural approach to quantifying performance is in carrying out a user study, comparing many pairs of images subject to a variety of nominal criteria. This is a sensible approach, but carries two limitations: The first issue with such an approach is that it is challenging to run a user study with a sufficiently large pool of subjects for a large enough set of images to claim that the results are broadly representative of overall algorithm performance. A second challenge, is that with each subsequent algorithm that is developed, it is necessary to repeat the study in contrasting new results with existing algorithms. Both of these issues are overcome in attempting to produce metrics that make predictions that are well aligned with observations from the user study. Both the efforts of Cădík et al. [20] and Yeganeh and Wang [31] present useful quantitative measures to this end. Framed in the context of the two axes of detail and realism, Cădík et al. [20] present an interesting study that casts existing efforts into this space subject to a large number of psychophysical measurements. It is interesting to note that existing algorithms are dispersed around the optimal result, with no single algorithm appearing to offer a great advantage over others. The Tone-Mapping Quality Index (TMQI) algorithm [31] has roots in the highly successful Structural Similarity index (SSIM) [32], and provides a measure of the quality of a tone-mapping result that appeals to structural integrity, and statistical correlations in image structure. A drawback of algorithmic approaches to assessment, is that the algorithms considered have a significant rank order correlation with human judgments, but not a precise alignment. For this reason, it's challenging to say that one specific algorithm is better than another based solely on a quantitative score. An additional challenge in assessment lies in the goals of HDR to LDR conversion, which as discussed may be variable. This may be one element that factors into subjectivity, since the association between individual opinions and nominal categorical judgments will no doubt differ across observers. This may also suggest that a fruitful avenue for further evaluation efforts may lie in investigating quantifiable indices that appeal to more specific categorical judgements. Nevertheless, such quantitative measures do provide a very useful adjunct to qualitative evaluation.

In the current study, algorithms have been compared subject to the recent TMQI algorithm presented by Yeganeh and Wang [31]. Given that the score for any given image may vary according to gamma adjustment, gamma adjustment has been performed for all values between 0.8 and 1.4 (at a precision of 0.01) to determine the optimal setting for each algorithm. In every case, a clear peak is present in this range, with an approximately normal distribution centered around the peak gamma adjustment value. Figs. 1, 3 and 6 include TMQI quality values in their annotations, and quantitative values with associated gamma values for each algorithm are presented in Tables 1 and 2.

It appears that as β is increased, the exponent for gamma adjustment applied to the ExpoBlend output that results in the optimal TMQI gradually converges to 1. Given the relationship

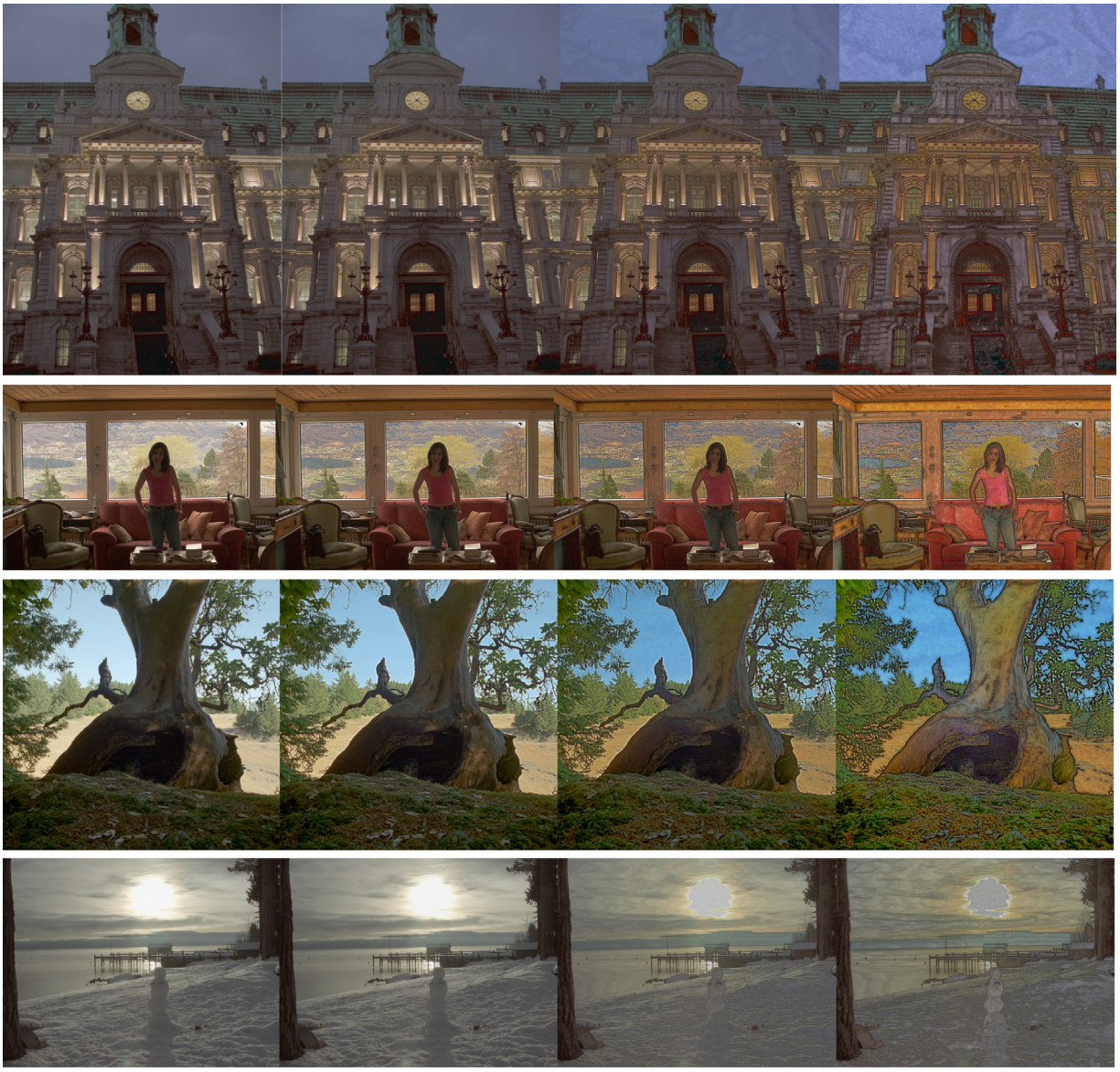

Fig. 8. Given existing HDR images, and taking between 23 and 30 bracketed 8-bit windows (in increments of 1/2 F-stop) between black and white images, the ExpoBlend algorithm is applied to the extracted intervals as a means of tone-mapping based on existing HDR format images. From left to right, $\beta = 5, 15, 30, 100$ respectively. Note the variation in detail that is possible as observed in Fig. 3.

between retaining local entropy across exposures, and preservation of structure present in the signal, this perhaps offers a sense of why this behavior is observed. It also appears that a strategy that implicitly seeks to optimize for structure across exposures scores high relative to the TMQI. That being said, the $\beta = 100$ example for the *memorial* image demonstrates that at the extreme of detail preservation one may observe a significant drop-off in quality score.

7. Discussion

In this paper, we have presented a novel algorithm for non-linear combination of separate exposures in the absence of camera response curves, exposure values or relative exposure values, irradiance, or any auxiliary information. This is shown to produce smooth low dynamic range results that capture the detail present across a wider dynamic range. In addition, it has been shown that the behavior of the approach presented provides a simple one parameter means of scaling between a smooth field that retains more coarse variation in illumination, or at the other extreme, exposing any detail that may be derived from the separate images. This has been demonstrated as a way of combining exposures when precise calibration is not possible, using a simple combination strategy that also may allow for more precise control over appearance of illumination or highlights in a blended image. The current implementation based on non-vectorized MATLAB code takes approximately 30 seconds to process a 1600×1200 image. Each stage of the algorithm is highly amenable to optimization and parallel processing implying the potential for a very fast implementation.

We have also demonstrated that this provides an additional means of tone-mapping HDR images to a lower dynamic range via simulated bracketing of the HDR image. In this second use case the ExpoBlend algorithm is well behaved both in the case that the relationship between intensities and irradiance has been linearized, and in the instance that an inverse gamma non-linearity is introduced.

The casting of exposures into the log domain appears to be essential to achieve some of the flexibility afforded by the algorithm, allowing for more selective control over the relative influence of

Fig. 9. An illustration of a number of individual exposures employed by the algorithm corresponding to the second row of Fig. 8. Note that local contrast and detail are widely spread across a large number of exposures, but the non-linear normalized entropy is sufficient to fuse these in a desirable manner ranging from realistic looking imagery that captures all the content in the image, to highly detailed stylized results in manipulating β.

Table 1
TMQI scores and associated exponents for gamma adjustment for the *memorial* image.

Algorithm	Gamma	TMQI
Durand and Dorsey [4]	1.26	0.8286
Fattal et al. [2]	1.27	0.8209
Reinhard et al. [3]	1.11	0.8658
Drago et al. [5]	1.15	0.8990
Reinhard and Devlin [6]	1.13	0.8503
Mantiuk et al. [7]	1.33	0.8079
Mantiuk et al. [8]	1.04	0.9393
ExpoBlend, $\beta=5$	1.18	0.9359
ExpoBlend, $\beta=15$	1.12	0.9322
ExpoBlend, $\beta=30$	1.06	0.9255
ExpoBlend, $\beta=100$	1.00	0.8782

Table 2
TMQI scores and associated exponents for gamma adjustment for Cădík et al.'s *lamp* image.

Algorithm	Gamma	TMQI
Cădík [21]	0.96	0.8043
Kim [22]	1.11	0.8645
Ferwerda et al. [23]	1.03	0.7835
Choudhury and Tumblin [24]	1.14	0.7979
Krawczyk et al. [25]	1.16	0.8681
Larson et al. [26]	1.07	0.8600
Kuang et al. [27]	1.11	0.7879
Goshtasby [10]	1.00	0.8232
Reinhard and Devlin [6]	1.20	0.7830
Ashikhmin [28]	1.10	0.8647
Drago et al. [5]	1.05	0.7995
Tumblin et al. [29]	1.04	0.8302
Duan et al. [30]	1.06	0.8360
ExpoBlend, $\beta=6$	1.15	0.8718
ExpoBlend, $\beta=24$	1.10	0.8782
ExpoBlend, $\beta=100$	1.04	0.8943

reflectance and illumination in determining the resultant fused image. Moreover, the softmax style fusion allows for implicit interpolation since all exposures contribute to any given pixel in the resulting image (subject to a weight), avoiding the need for explicit spatial interpolation. This also presents the interesting possibility of adapting the local filter aperture over which information is quantified. In particular, in the instance that all exposures elicit a low entropy value, the region of support might be grown subject to a minimum threshold to allow for both locality and continuity as discussed.

An additional consideration of importance, is the realm of the parameter space where the model is well behaved. Generally this range is relatively wide, but as the number of exposures available decreases down to a small number, the sensitivity of this variable increases, and the possibility of introducing artifacts, such as banding on uniform surfaces, or localized halo effects becomes

more significant. This again appeals to the minimum required support for determining the contribution of different exposures. This is seen in one example, in the moon area (Fig. 8, bottom row) for larger parameter values, in which the moon becomes darker as β is increased. In practice, interpolation, or a multi-scale strategy would help with this, but at the expense of the locality that provides strong contrast in cases such as the window region in the *memorial* image. In our observations, with high quality imagery, this is less of an issue since even for relatively homogeneous regions, there is some meaningful signal that is most prevalent at desirable exposures. For lower quality imagery from webcams, or camera phones this consideration gains additional importance.

The equivalence of the two latter stages of normalization to the softmax activation function that appears in work on neural networks hints at some of the generality of the proposal. The information seeking impetus of the algorithm implies that this might also be applied to images with varying depth of field to optimize focus, multi-spectral or multi-sensory data, and other similar scenarios. It also implies amenability to a learning strategy, which might employ local summary statistics associated with regions of separate exposures towards optimizing output subject to a desired principled objective function.

As a whole, the current work presents a useful adjunct to existing strategies for instances for which precise radiometric calibration is not possible, as a quick and crude way of combining distinct exposures while also arguably producing superior results in some instances to alternative strategies. One particular strength of the approach appears to be preservation of detail among the brightest areas of a scene carrying challenging lighting conditions. Moreover, the behavior of the algorithm may be modified in manipulating a single parameter to adjust the output to optimize for contrast, photo-realism, or emphasis of detailed structure.

Acknowledgments

The authors wish to thank the reviewers for their valuable comments and suggestions in improving the presentation of this work. The authors gratefully acknowledge the financial support of NSERC and the University of Manitoba.

References

[1] Bruce NDB. Non-linear normalized entropy based exposure blending. In: Proceedings of graphics interface 2013. CRC Press; 2013. p. 37–44.
[2] Fattal R, Lischinski D, Werman M. Gradient domain high dynamic range compression. In: Proceedings of SIGGRAPH 2002. ACM Press; 2002, p. 249–56. http://doi.acm.org/10.1145/566570.566573.
[3] Reinhard E, Stark M, Shirley P, Ferwerda J. Photographic tone reproduction for digital images. In: Proceedings of SIGGRAPH 2002. ACM Press; 2002. p. 267–76. http://dx.doi.org/10.1145/566570.566575.
[4] Durand F, Dorsey J. Fast bilateral filtering for the display of high-dynamic-range images. In: Proceedings of SIGGRAPH 2002. ACM Press; 2002. p. 257–66. http://doi.acm.org/10.1145/566570.566574.
[5] Drago F, Myszkowski K, Annen T, Chiba N. Adaptive logarithmic mapping for displaying high contrast scenes. Comput Graph Forum 2003;22(3):419–26.
[6] Reinhard E, Devlin K. Dynamic range reduction inspired by photoreceptor physiology. IEEE Trans Vis Comput Graph 2005;11(1):13–24.
[7] Mantiuk R, Myszkowski K, Seidel HP. A perceptual framework for contrast processing of high dynamic range images. ACM Trans Appl Percept 2006;3(3):286–308.
[8] Mantiuk R, Daly S, Kerofsky L. Display adaptive tone mapping. ACM Trans Graph 2008;27(3):68:1–10, http://dx.doi.org/10.1145/1360612.1360667.
[9] Mertens T, Kautz J, Reeth FV. Exposure fusion: a simple and practical alternative to high dynamic range photography. Comput Graph Forum 2009;28(1):161–71.
[10] Goshtasby AA. Fusion of multi-exposure images. Image Vis Comput 2005;23(6):611–8, http://dx.doi.org/10.1016/j.imavis.2005.02.004.
[11] van Oppenheim A, Schafer R, Stockham Jr. T. Nonlinear filtering of multiplied and convolved signals. IEEE Trans Audio Electroacoust 1968;16(3):437–66.
[12] Tappen MF, Freeman WT, Adelson EH. Recovering intrinsic images from a single image. IEEE Trans Pattern Anal Mach Intell 2005;27(9):1459–72.
[13] Bridle JS. Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters. In: Advances in neural information processing systems 2. Morgan Kaufmann Publishers Inc.; 1990. p. 211–7.
[14] Ogden JM, Adelson EH, Bergen JR, Burt PJ. Pyramid-based computer graphics. RCA Eng 1985;30(5):4–15.
[15] Burt PJ, Hanna K, Kolczynski R. Enhanced image capture through fusion. Proceedings of the workshop on augmented visual display research, NASA—Ames Research Center 1993;1:207–24.
[16] Debevec PE, Malik J. Recovering high dynamic range radiance maps from photographs. In: Proceedings of SIGGRAPH 1997. ACM Press; 1997. p. 369–78. http://dx.doi.org/10.1145/258734.258884.
[17] Cadik M. Evaluation of tone mapping operators. 2013. ⟨http://dcgi.felk.cvut.cz/home/cadikm/tmo/⟩.
[18] Mantiuk R, Krawczyk G, Mantiuk R, Seidel HP. High dynamic range imaging pipeline: perception-motivated representation of visual content. In: Human Vision and Electronic Imaging XII, vol. 6492; 2007. p. 649212. http://dx.doi.org/10.1117/12.713526.
[19] Cadik M, Wimmer M, Neumann L, Artusi A. Image attributes and quality for evaluation of tone mapping operators. In: Proceedings of the 14th Pacific conference on computer graphics and applications. Taipei, Taiwan: National Taiwan University Press; 2006. p. 35–44.
[20] Cadik M, Wimmer M, Neumann L, Artusi A. Evaluation of HDR tone mapping methods using essential perceptual attributes. Comput Graph 2008;32:330–49.
[21] Cadik M. Perception motivated hybrid approach to tone mapping. In: WSCG (full papers). 2007. p. 129–36.
[22] Kim MH. High-fidelity colour reproduction for high-dynamic-range imaging. Ph.D. thesis; University College London; 2010.
[23] Ferwerda JA, Pattanaik SN, Shirley P, Greenberg DP. A model of visual adaptation for realistic image synthesis. In: Proceedings of SIGGRAPH 1996. ACM; 1996. p. 249–58.
[24] Choudhury P, Tumblin J. The trilateral filter for high contrast images and meshes. In: ACM SIGGRAPH 2005 courses. New York, NY, USA: ACM; 2005. http://doi.acm.org/10.1145/1198555.1198565.
[25] Krawczyk G, Myszkowski K, Seidel HP. Lightness perception in tone reproduction for high dynamic range images. Comput Graph Forum 2005;24(3):635–45.
[26] Larson GW, Rushmeier H, Piatko C. A visibility matching tone reproduction operator for high dynamic range scenes. IEEE Trans Visual Comput Graph 1997;3(4):291–306.
[27] Kuang J, Johnson GM, Fairchild MD. iCAM06: a refined image appearance model for HDR image rendering. J Vis Comm Image Represent 2007;18(5):406–14.
[28] Ashikhmin M. A tone mapping algorithm for high contrast images. In: Proceedings of the 13th eurographics workshop on rendering. Eurographics Association; 2002. p. 145–56.
[29] Tumblin J, Hodgins JK, Guenter BK. Two methods for display of high contrast images. ACM Trans Graph (TOG) 1999;18(1):56–94.
[30] Duan J, Qiu G, Chen M. Comprehensive fast tone mapping for high dynamic range image visualization. In: Proceedings of pacific graphics 2005; 2005.
[31] Yeganeh H, Wang Z. Objective quality assessment of tone-mapped images. IEEE Trans Image Process 2013;22(2):657–67, http://dx.doi.org/10.1109/TIP.2012.2221725.
[32] Wang Z, Bovik A, Sheikh H, Simoncelli E. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 2004;13(4):600–12, http://dx.doi.org/10.1109/TIP.2003.819861.