RANGE-ENHANCED ACTIVE FOREGROUND EXTRACTION

Asla Sá, Marcelo B. Vieira, Paulo Cezar Carvalho and Luiz Velho

IMPA - Instituto de Matemática Pura e Aplicada

ABSTRACT

We describe a new technique that uses active scene illumination to perform foreground-background segmentation and recover partial HDR information. We explore the fact that relative tones can be recovered by varying illumination intensity, without knowing the camera response function. In our approach, we illuminate the scene with an uncalibrated projector and capture two images of the scene under different illumination conditions. By taking advantage of the fact that the projector can be set up to illuminate only the foreground, we are able to distinguish the foreground from the background. The output of our system is a segmentation mask, together with a image with additional tonal information for the foreground pixels. As an application, we show how to produce spatially variant tone mapped images, where background and foreground receive different treatments. The segmentation and the visualization algorithms are implemented in real-time, and can be used to produce range-enhanced video sequences.

1. INTRODUCTION

In this paper, we describe a method to acquire HDR information for the foreground of a scene, which is automatically segmented from the background. In order to do that, we use a controlled (but uncalibrated) light source, such as a video projector or a flash, to capture images of the scene under different illumination conditions (modulating light intensity). By taking advantage of the fact that the projector light can be set up to illuminate only the foreground, we are able to segment the image in illuminated/non-illuminated regions, corresponding to the foreground and the background, respectively. In addition, the regions modulated by the projector light have an enhanced tonal information. We argue that this amounts to an enhanced dynamic range (HDR) image recovery, in which relative, rather than absolute, tone values are employed.

The recovered relative tone values can be used to produce new images of the scene by applying a spatially-variant tone mapping operator, where background and foreground are treated in a different manner. The entire process (illumination, capture, segmentation, tone mapping) can be done in real-time using off-the-shelf equipment and can be used to produce real-time tone-enhanced video.

The paper is organized as follows. In section 2 we review the main concepts related to HDR images and tone-mapping operators. In section 3 we discuss the use of relative tone values and show that, for many applications, they can be used instead of absolute tones. In section 4 we show how to perform foreground-background segmentation with the aid of active illumination and how to build a spatially-variant tone mapping operator. In section 5 we detail our implementation and show some results. Final comments and future research directions appear in section 6.

2. PRELIMINARIES

Most of the available imaging systems and displays have a very limited range, typically two orders of magnitude. The need to overcome these limitations has led to the study of the recovery of High Dynamic Range (HDR) radiance maps from Low Dynamic Range (LDR) images. The standard technique used to recover HDR images is to acquire some differently exposed images of a scene and apply the camera response curve to obtain its radiance values [3, 4]. Below, we discuss this process in more detail.

To form an image, an imaging sensor converts real irradiance values of a scene into digital brightness values. The imaging sensor's behaviour is described by a characteristic response function $f : [E_{min}, E_{max}] \rightarrow [0, 1]$ that maps irradiance to brightness values; where E_{min} and E_{max} are respectively the minimum and the maximum values measurable by the sensor. It is reasonable to assume f to be monotonically increasing; thus, its inverse f^{-1} is well defined.

A set of N differently exposed pictures of a scene gives a set of values $(B^k_{d_{ij}})$ which can be used to recover f and the actual scene irradiance. Considering that k is the index on exposure times, and $E^k = B_{w_{ij}} \Delta t_k$ is the exposure of image at pixel ij then $B_{d_{ij}}^k = f(B_{w_{ij}} \Delta t_k)$ where $k = 1..N$. To recover f some assumptions are imposed on the function, for example, the use of parameterized models such as a gamma curve, or continuity and monotonicity restrictions. To recover the actual irradiance values, f^{-1} is applied to the correspondent pixels brightness values $B_{w_{ij}}$, that is,

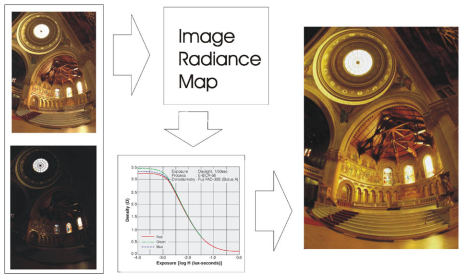

Fig. 1. Differently exposed pictures are processed to compose a HDR Radiance Map and a TMO (in this case, a film response curve) is applied to visualize it.

$B_{w_{ij}} = \frac{f^{-1}(B_{d_{ij}}^{k})}{\Delta t_k}$. If the pixel is almost over or underexposed, a lower weight is given to it, augmenting the influence of the middle of the f curve, where sensors (and films) are well behaved. In Figure 1 we show the pipeline of HDR recovery from photographs.

Instead of varying the exposure, one could think of an alternative way to acquire HDR images, in which the exposure is kept constant, while changing the intensity of the illumination, in order to record lighter or darker areas of the scene. This is not usually done, since the illumination intensity varies irregularly through the scene, making it difficult to retrieve absolute tone values. However, the concept of relative tone values introduced here makes possible the use of active illumination in recovering relative tones to enhance tonal resolution on the illuminated area.

All discussion above assumes that one knows precisely the correspondence between pixels in different images. This is trivial for static scenes captured by cameras mounted on a tripod. For moving scenes, one has to determine pixel correspondence before applying the method outlined above. However, Grossberg and Nayar [4] have proven that the image histogram contains all information needed to recover f. The use of histograms has many advantages: they are less sensitive to noise, all the information present in the image is used (instead of just a subset of pixels), and the correspondence restriction is relaxed.

Usually, the HDR data is visualized on a display of limited range. In this case, the image range must be adjusted to display's lower range. The idea is to arrange brightness maintaining color chrominance values. This problem is known as the Tone Mapping Operator (TMO) problem, in which the goal is to devise tone maps that are capable of producing good perceptual results. Several heuristics have been proposed to compress the range of radiances of HDR images, and the comparison between them is mainly perceptual, see [1].

The simplest way to reduce the range of an HDR is a linear map, but the resulting image is frequently not perceptualy pleasant. In [2], Larson et al. propose the histogram adjustment heuristic, that is inspired on the fact that luminance levels typically occurs in clusters, rather than being uniformly distributed throughout the range. The brightness values are given by $B_d = P(B_w)(B_{d_{max}} - B_{d_{min}}) + B_{d_{min}}$, where $P(B_w) = \frac{[\sum_{b_i < B_w} h(b_i)]}{[\sum_{b_i < B_{w_{max}}} h(b_i)]}$, where h is the tone frequency histogram.

3. ABSOLUTE VS. RELATIVE TONE VALUES

In this section, we introduce the conceptual difference between absolute and relative tone values in the context of HDR data recovery. Of particular interest to our discussion is the paper [6] where the authors prove the simple summation theorem:

Theorem 1 (Simple Summation) *The sum of a set of images of a scene taken at different exposures includes all the information in the individual exposures.*

This means that, if one knows the exposure times used to obtain the images and the response curve of the camera, one can recover the irradiance values and the individual images from their sum. In [6], the authors use this result to optimize the camera parameters. In this paper, we pose the following question: *what can be recovered from the sum of the images, if one does not know neither the exposures nor the response curve?* Surprisingly, the answer is: many things!

We define *relative tones* q as the values present in the summation image, while *absolute tones* B_w are the real irradiances. The relative q values are unique indexes to real irradiance values. Thus, with the response funcion f and the exposure camera parameters in hand we can generate a look-up table mapping q to B_w values, i.e., $F_{f,\Delta_t} : [0, 2] \rightarrow [E_{min}, E_{max}]$. In Figure 2 we illustrate the relation between the recovered quantization levels q and the absolute tone values B_w. We observe that this mapping F is 1-1.

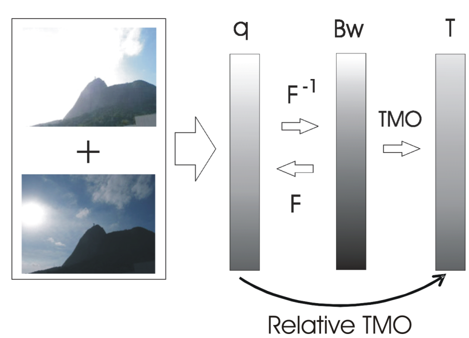

Fig. 2. Absolute vs. Relative tone values.

Usually, TMO's are described in terms of absolute tones. However, since there is a 1-1 mapping between relative and absolute tone values, we conclude that we can propose TMO's to be applied directly to the relative tones.

4. ACTIVE RANGE-ENHANCEMENT

We observe that controlling light intensity, while keeping all other camera parameters fixed, we change the pixel irradiance $B_{w_{ij}}$ and hence its exposure value. Our goal is to obtain an image with augmented tonal resolution through *active illumination*.

In order to recover absolute tones B_w from varying illumination, one would need not only to calibrate the light sources but also to model the light propagation and interaction in the scene. However, relative tones are recovered naturally in regions where the tonal partial ordering is preserved.

The fact that illumination mostly affects the foreground pixels can be explored to perform foreground-background segmentation. Below, we discuss the details of the segmentation process, and of the acquisition and visualization of relative tones for the foreground.

4.1. Active Foreground Extraction

Suppose we have an image $\mathrm{I}^k = \{\mathrm{p}_{ij}^k\}$ acquired from a scene where the foreground was uniformly illuminated with a light intensity $\mathrm{g}^k \in [0,1]$. We assume that the image is composed by 2 disjoint subsets, i.e., $\mathrm{I}^k = \mathrm{O}^k \cup \mathrm{S}^k$ where O^k is the foreground illuminated region and S^k is the background.

In order to extract the foreground, we acquire another image $\mathrm{I}^{k+1} = \{\mathrm{p}_{ij}^{k+1}\}$ where the foreground is illuminated with light intensity $\mathrm{g}^{k+1} \in [0,1]$, with $\mathrm{g}^k \neq \mathrm{g}^{k+1}$. We assume that background objects and illumination do not change significantly between the consecutive shots.

The difference in lighting is sufficient to detect objects affected by the active light. For the foreground-extraction phase we use the luminances $L(\mathrm{p}_{ij}^k)$.

Given a sequence of different illuminated images, we define the foreground as $\mathrm{O}^k = \{\mathrm{p}_{ij}^k \mid |L(\mathrm{p}_{ij}^k) - L(\mathrm{p}_{ij}^{k+1})| \geq l_{min}\}$, i.e. the set of points $\{\mathrm{p}_{ij}\}$ having absolute difference of luminance in both images greater than or equal to a threshold l_{min}.

4.2. Spatially Variant Tone Mapping

Using active illumination, the foreground objects not only can be extracted but also can be range-enhanced using relative tone mapping. We define the *enhanced range luminance* as $R(\mathrm{p}_{ij}^k) = L(\mathrm{p}_{ij}^k) + L(\mathrm{p}_{ij}^{k+1})$, $R(\mathrm{p}_{ij}^k) \in [0,2]$ based on the Simple Summation Theorem.

To visualize the foreground tones we directly apply Larson's histogram adjustment $\mathrm{TMO}_{sv} : [0,2] \to [0,1]$ to the set $\{R(\mathrm{p}_{ij}^k) \mid \mathrm{p}_{ij}^k \in \mathrm{O}^k\}$. Cromaticities can be linearly combined to avoid problems in over or under-exposed areas. We use the average of both cromaticities in our experiments.

Note that the background cannot be range-enhanced since $L(\mathrm{p}_{ij}^k) \approx L(\mathrm{p}_{ij}^{k+1})$ in this region. It is worthy to point out, however, that the background can also be especially processed. More generally, multiple regions having different range characteristics can be processed separately, inducing a *spatially variant tone mapping*.

The combination of the different tone mapped regions can be done by simple image composition. We now assume the case where the foreground pixels p^k are range-enhanced and the background pixels q^k remain unchanged. Let M^k be a mask image where 1 is assigned to the foreground pixels and 0 to the background. The composed image pixels can be defined as $\{C(\mathrm{c}_{ij}^k) = C(\mathrm{p}_{ij}^k) \cdot C(\mathrm{m}_{ij}^k) + C(\mathrm{q}_{ij}^k) \cdot [1 - C(\mathrm{m}_{ij}^k)]\}$ where $\mathrm{m}_{ij}^k \in \mathrm{M}^k$ and $C(\mathrm{p})$ is the color at pixel p. We apply a low-pass filter to image M^k to avoid the perception of discontinuities in transition regions.

5. RANGE-ENHANCED VIDEO

The results of Sections 4 and 3 can be applied to obtain a range-enhanced video stream. We propose a system composed by a video camera synchronized with a projector. A video signal, where each field has constant gray color values g^1 and g^2 respectively ($\mathrm{g}^1 \neq \mathrm{g}^2$) is projected onto the scene. This signal is connected to the camera *gen*-lock pin, which guarantees projection/capture synchronization.

In this way, the objects subject to the projector light will be lit with intensity g^1 and g^2 alternately. Each light exposure lasts for $1/59.54s$ using the NTSC standard. Range-enhanced video can be produced once the camera is capturing the active illuminated scene.

The fields I^1 and I^2 of each captured frame represent the same object illuminated differently. As shown in Section 4, this is sufficient for doing an active foreground extraction and range-enhancement.

Input video images are in the Yuv color space. Thus, the processing is performed using the luminance defined as $L(\mathrm{p}_{ij}^k) = \mathrm{Y}(\mathrm{p}_{ij}^k)$, where Y is the luminance channel. In this scheme, the spatially variant range-enhacement proposed in Section 4 can be directly applied to any two consecutive fields. This achieves a range-enhanced output video stream with the same input framerate, as shown in Fig. 3.

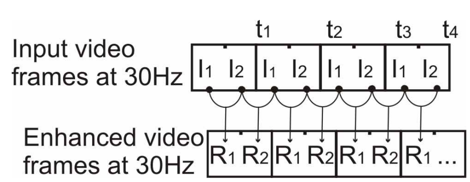

Fig. 3. Two consecutive input fields result in one frame.

We assume that the framerate is high compared to the object's movement and, as a consequence, the effects of moving objects or camera movement are small between a pair of fields.

Figure 4 shows two consecutive video fields with different illumination, the filtered foreground mask and its resulting tone enhancement. The gray values used are $g^1 = 0.35$ and $g^2 = 1$. The luminance threshold L_{min} used is 0.08. One can notice that our method has some trouble segmenting low-reflectance objects, such as hair. However, the resulting tone-enhanced image is still quite satisfactory. Shadows may also be a problem, but can be treated in many ways (e.g., we can use more than one light source, or position the light near the camera). Other results can be seen at `http://www.impa.br/~asla/ahdr`.

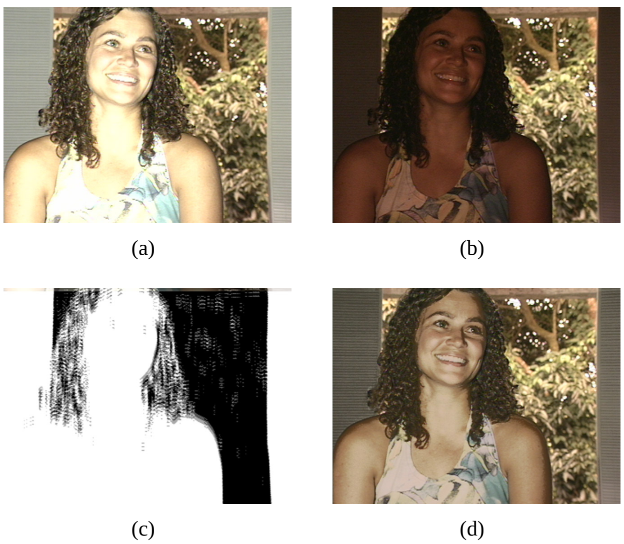

Fig. 4. Images (a) and (b) are the input fields, (c) is the foreground mask, (d) shows the range-enhanced foreground.

6. CONCLUSIONS

We have presented an active illumination technique that produces foreground-background segmentation of a scene and also generates enhanced range tonal information of the foreground region. This technique is made possible by the concept of relative tone values introduced in this paper. We remark that relative tones is a key concept in HDR theory and has many other applications, which we intend to exploit in future work.

We have also developed a system for range-enhanced video. The acquisition device is composed of synchronized camera and projector. It is built with off-the-shelf NTSC equipment. This system has many advantages, such as good cost performance, compatibility and various options of distribution channels. The data processing of our system can be easily incorporated into the pipeline of a video production environment.

Although our implementation has been done in real time for video, the same idea could be used in digital cameras programming flash. There are many recent works [10, 11] that explore the use of programmable flash to enhance image quality, but they do not introduce HDR concepts. Our work gives a contribution to this new area of computational photography.

7. REFERENCES

[1] F. Drago, W.L. Martens, K. Myszkowski and H.P. Seidel, *Perceptual Evaluation of Tone Mapping Operators with Regard to Similarity and Preference*, Tech. Report MPI-I-2202-4-002 (2002).

[2] G.W. Larson, H. Rushmeier and C. Piatko. *A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes*, IEEE Trans. on Vis. and Comp. Graph., 3(4): 291-306 (1997)

[3] P.E. Debevec and J. Malik, *Recovering High Dynamic Range Radiance Maps from Photographs*, Proc. ACM SIGGRAPH 97 - pp.369-378 (1997).

[4] M.D. Grossberg and S.K. Nayar, *Determining the Camera Response from Images: What Is Knowable?*. IEEE Trans.PAMI vol.25 no.11 (2003).

[5] M.D. Grossberg and S.K. Nayar, *Modeling the Space of Camera Response Functions*. IEEE Trans.PAMI vol.26 no.10 (2004).

[6] M.D. Grossberg and S.K. Nayar, *High Dynamic Range from Multiple Images: Which Exposures to Combine?*. In Proc. ICCV Workshop CPMCV (2003).

[7] S.B. Kang, M. Uyttendaele, S. Winder and R. Szeliski, *High Dynamic Range Vdeo*, Proc. ACM SIGGRAPH 03 - pp.319-325 (2003).

[8] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh and A. Vorozcovs, *High Dynamic Range Display Systems*, Proc. ACM SIGGRAPH 04 - pp.?-? (2004).

[9] M. B. Vieira, L. Velho, A. Sá, and P. Carvalho, *Real-Time 3D Video*, Visual Proc. of SIGGRAPH. ACM, 2004.

[10] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe and K. Toyama, *Digital photography with flash and no-flash image pairs*, ACM Trans. on Graphics 23(3): 664-672 (2004).

[11] E. Eisemann and F. Durand, *Flash photography enhancement via intrinsic relighting*, ACM Trans. on Graphics 23(3): 673-678 (2004).